

# 場所表現の地理的曖昧性を解消するための質問内容生成

清水美緒奈 林純子 久田祥平 若宮翔子 荒牧英治 大内啓樹

奈良先端科学技術大学院大学

shimizu.miona.sl1@naist.ac.jp

{hayashi.junko.hh5, s-hisada, wakamiya, aramaki, hiroki.ouchi}@is.naist.jp

## 概要

場所表現の地理的曖昧性は、ジオコーディングにおいて重要な課題である。従来の研究では、場所表現が登場する文脈情報の有用性が示されている一方で、文脈情報を考慮しても位置が特定できない場合も少なくない。そこで、本研究では、曖昧な場所表現の位置を特定するためのユーザへの質問内容生成を目的として、複数の場所候補を絞り込む上で手がかりとなるランドマークの抽出を試みた。実験では、GPT-4oを使用し、人手で作成した評価データを用いてその有用性を評価した。結果として、モデルは中程度の精度を示したが、カテゴリの曖昧さや過剰な推測に起因する正解ランドマークの見逃しが多い傾向にあることが分かった。

## 1 はじめに

地名や施設名など、実世界の特定の場所を指す言語表現（場所表現）の緯度経度を推定することは非常に重要である。このようなタスクをジオコーディングと呼び、応用先は、旅行記や感染症の積極的疫学調査、110番への緊急通報など多岐にわたる [1]。

場所表現の位置推定の手がかりとなる情報は、場所表現が登場する文脈である。しかし、文脈を考慮しても位置が一意に特定できない場合も少なくない。「日本橋にいます。」という例では、「日本橋」という同一の地名が東京と大阪の両方に存在するため、この文脈だけではどちらの地域の「日本橋」か特定できない。このような例では、ユーザにどちらの「日本橋」を指しているかを質問することができれば、その場所を特定することができる。

そこで本研究では、曖昧な場所表現の位置を特定するためのユーザへの質問生成を想定し、その質問内容の生成を試みる。より具体的には、ある曖昧な場所表現がある場合、いくつかの候補を列挙し、その候補を絞り込む上で手がかりとなる周辺ランド

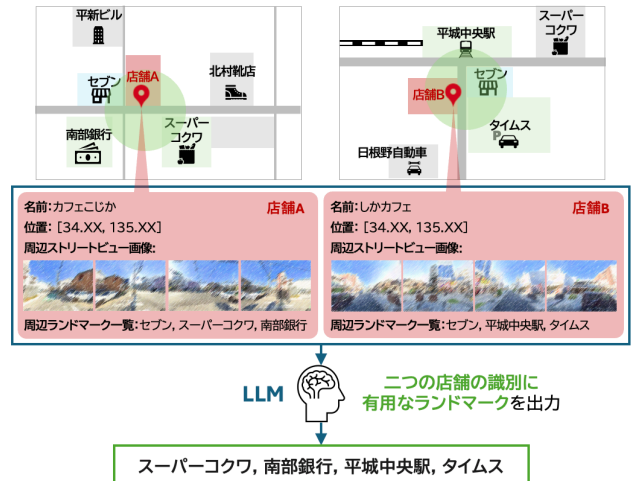


図1 識別に有用なランドマーク抽出モデルの入出力の例

マークの抽出を試みる。このような対話を想定した位置の特定技術は、既存のジオコーディングタスクの範疇を超えた新たな技術であり、非常に新規性が高く、応用範囲も広い。

Large Language Model (以下、LLM) を用いた実験を行い、評価データセットにもとづいて抽出したランドマークが、どのくらいの割合で正解の候補を絞り込むことができるかを評価した。その結果として、モデルは中程度の精度を示したが、カテゴリの曖昧さや過剰な推測に起因する正解ランドマークの見逃しが多い傾向にあることが分かった。

## 2 タスク設定

### 2.1 想定シナリオ

ユーザが曖昧性のある場所表現を記述したことを想定する。例えば「カフェに行ってきました。」という文における「カフェ」が場所表現に該当する。実用上では、ある場所表現についてその位置の候補の絞り込みも行う必要があるが、本研究ではタスク設定を簡略化し、二つの候補店舗まで絞り込んだ状況を想定する。つまり、ある場所表現について、



図2 店舗A, B周辺のストリートビュー画像例 (左から順に東, 南, 西, 北方向)

店舗Aと店舗Bまで絞り込んだ状況を想定する。

正解の位置を特定するため、「そのカフェの近くには、{landmark}はありますか?」のような質問を生成することを想定する。ここで{landmark}には実際のランドマーク名を入れる。重要なのは、二つの候補店舗を識別可能なランドマークを選定することである。例えば「セブンイレブン」を例にとると、店舗Aと店舗Bのどちらか一方の近くにだけ「セブンイレブン」があれば候補店舗を絞り込めるが、どちらの近くにも「セブンイレブン」があるなら絞り込めない。したがって、質問に加えることで候補店舗を絞り込めるランドマークを、機械学習モデルによって選択するタスクとして設計する。

## 2.2 タスクの入出力

図1にタスクの概要を示す。モデルには以下の三つの情報を入力として与える。

- 入力1: 各店舗の名前・位置 (緯度経度)
- 入力2: 各店舗周辺の四方位 (東西南北) のストリートビュー画像 (図2参照)
- 入力3: 各店舗周辺のランドマーク一覧

入力2では、GoogleのStreetViewStaticAPI<sup>1)</sup>を用いて各店舗周辺の四方位の画像を取得した。また、入力3では、各店舗から距離の近い10軒のランドマークの地物名を候補としてリスト化し、さらにそれらの地物名に対応する抽象度の高いカテゴリ (amenity, shop, public\_transport など) も候補に加えた。

モデルはこれらの入力情報をもとに、ランドマーク一覧から識別に有用なランドマークを出力する。出力結果は、人手で作成した評価データを用いて評価した。

1) <https://developers.google.com/maps/documentation/streetview/overview?hl=ja>

## 3 評価データ構築

**候補店舗の選定** 奈良市内のカフェを対象とし、二つのカフェ (店舗A, 店舗B) を識別するために有用なランドマークを抽出することを目的とした。具体的には、OpenStreetMap (OSM)<sup>2)</sup>のデータを用いて、奈良市内の「amenity=cafe」タグが付与された159軒のカフェを抽出した。そして、抽出したカフェ集合の中から無作為に20ペアを作成した。

**周辺ランドマークの絞り込み** まず、2.2節の入力3と同様に、店舗Aと店舗Bのそれぞれについて、距離の近い10軒のランドマークを抽出した。次に、抽出した全20軒のランドマークの中から、各店舗周辺のストリートビュー画像をもとに、人間が視認できるものだけに絞り込んだ。

**ランドマークの正解候補を選出** 絞り込んだランドマークの中から、図3に示すように、各ランドマーク候補について地物名とカテゴリの二段階で表現し、抽象度を考慮しつつ正解ランドマークを選出した。具体的には、まず、ランドマーク候補の地物名同士を店舗A, Bで比較し、一致するものがない場合は、そのランドマークの地物名を正解候補として選出した。次に、同様の方法でカテゴリ同士を比較し、一致するものがない場合は、そのランドマークのカテゴリについても正解候補とした。

**正解ランドマークの最終決定** 実際に正解候補のランドマークを人間に提示した際に、店舗Aと店舗Bを識別できるかを評価し、その評価に基づいて最終的な正解ランドマークを決定した。具体的には、各店舗周辺のストリートビュー画像とともに「その店舗は{landmark}の近くにありました」という文を3人の評価者に提示した。評価者には、その文が

2) <https://www.openstreetmap.org/>

## 視認性で絞られた ランドマーク候補



図3 抽象度を考慮した正解ランドマーク選定の例

店舗Aと店舗Bのどちらを指しているかを回答してもらい、評価者の過半数が正しく回答できた候補のみを正解ランドマークとして採用した。

## 4 実験

### 4.1 実験設定

モデルとしてGPT-4o<sup>3)</sup>を使用した。GPT-4oへの入力プロンプトを図4に示す。「タスク概要」では、友人から「以前勧められたカフェに行ったが、その名前を忘れてしまった」と質問されたユーザが、奈良市内のどのカフェかを推測するために、友人にカフェ周辺のランドマークの有無を尋ねるという状況を設定している。「提供情報」には、各カフェの名前、緯度経度とストリートビュー画像<sup>4)</sup>に関する説明、そして各カフェ周辺のランドマーク候補の一覧<sup>5)</sup>が含まれる。どちらのカフェにも識別に有用なランドマークがない場合の選択肢として「-1」を追加した。「タスク内容」では、各カフェ周辺のストリートビュー画像を比較し、カフェを識別する手がかりとなるランドマークの番号を回答するように指示した。また、適切なランドマークが見つからない場合には「-1」を選択するよう求めた。

### 4.2 実験結果

評価データに対する実験結果を表1に示す<sup>6)</sup>。Recall値は0.5未満と低い傾向を示しており、モデ

**【タスク概要】**  
あなたの友達は、あなたが以前勧めたカフェに行きました。1ヶ月後、友達から電話で「勧めてもらったカフェに行ってきたんだけど、カフェの名前を忘れたので教えてほしい」と尋ねられました。しかしあなたはその友達に、2軒のカフェを勧めており、どちらのカフェに行ったのかわかりません。そこであなたは、以前撮影したそれぞれのカフェの周辺画像をもとに、そこに写り込んでいるランドマークの有無を聞けば、どちらのカフェに友達が行ったか分かったと思います。

**【提供情報】**

- カフェA
  - ・名前: カフェバルドー
  - ・位置: [34.6866135, 135.8262264]
- カフェB
  - ・名前: シャララ
  - ・位置: [34.6817489, 135.8210773]
- ストリートビュー画像  
各カフェ(AとB)周辺の4方位(東西南北)のストリートビュー画像を1枚ずつ添付しています。1~4枚目はカフェA周辺、5~8枚目はカフェB周辺の画像です。
- 周辺ランドマーク一覧
  - ・カフェA周辺のランドマーク
  - A-1 カクレミノ屋
  - ：
  - A-17 building(※buildingとは、家屋、工場、廃墟といった、様々な種類の建物を指します。)
  - ・カフェB周辺のランドマーク
  - B-1 酒楽
  - ：
  - B-13 amenity(※amenityとは、飲食店、駐車場、銀行、薬局、刑務所、学校など、来訪者や住人にとって便利で重要な場所を指します。)
  - 1 手がかりのランドマークなし

**【タスク内容】**  
2軒のカフェの周辺画像を見比べ、どちらのカフェかを区別するための手がかりとなるランドマークを「周辺ランドマーク一覧」から選択し、その番号を回答して下さい(複数選択可)。もし、手がかりとなるランドマークが見当たらない場合は、「-1」を選んでください。

図4 GPT-4oへの入力プロンプト例

ルにより出力されたランドマークは正解ランドマークを十分に網羅できていないことが示唆された。また、モデル全体の評価として、F1値も中程度にとどまっており、二つのカフェを識別する上で有用なランドマークの抽出性能は限定的である。

### 4.3 考察

Recallの低下を招く要因として、モデルに選択されなかった正解ランドマークの特徴には、主に以下の二つのケースが見られる。

3) <https://platform.openai.com/docs/models#gpt-4o>  
 4) ストリートビュー画像は、プロンプトに添付する形で各カフェ周辺の四方位の画像を一枚ずつ挿入した。  
 5) 各カテゴリ名にはカテゴリの定義文を与えた。  
 6) 同一の実験を5回試行し、その平均値と標準偏差を示している



表1 評価データに対するモデルの実験結果

Precision	Recall	F1
54.2 ( $\pm 0.410$ )	47.8 ( $\pm 0.379$ )	48.0 ( $\pm 0.360$ )

一つ目は、カテゴリ間の曖昧さに起因して正しい選択肢が見逃されたケースである。選択肢として与えられたカテゴリの情報が不十分であったため、カテゴリ間の区別が曖昧になり、その結果、正しい選択肢が選択されなかった可能性が考えられる。特に、「amenity」と「shop」の二つのカテゴリにおいては、どちらも日常的に利用される施設であり、機能や外見において重なる部分が多いため、モデルがどちらに分類すべきかを明確に判断できなかったと推測できる。このように、類似した特徴を持つカテゴリが同じ選択肢に含まれる場合は、正しいランドマークを見逃すリスクが高くなるため、カテゴリ間の境界をより明確に定義する必要がある。

二つ目は、「-1」（識別に有用なランドマークがない場合）が正解であるにも関わらず、モデルが過剰な推測を行ったケースである。二つのカフェ周辺ともに目立つ特徴が少ない状況で、モデルが他の地物に関連性を見出し、過剰に推測してしまうことが原因であると考えられる。例えば、周囲の地物が同じような特徴を持っている場合、モデルはその視覚的な類似性を過大評価し、関連性が薄い他の選択肢を選んでしまうことがある。特に、「-1」が選ばれるべき状況であっても、目立つランドマークが少ない場合、モデルは未知の特徴を過剰に引き寄せ、誤った選択をする傾向にあった。

これらを踏まえ、モデルがより適切に選択肢を絞り込むためには、モデルの判断基準の明確化やプロンプト設計の改善が必要であると考えられる。

## 5 関連研究

本研究と関係の深いジオコーディング研究では、機械学習を用いた手法が多く提案されている [2, 3]. 具体的には、地名辞典や地理空間における特徴表現を活用して場所表現の地理的分布を表現し、それらを言語的特徴と統合することにより、正確な位置推定を実現した。これらの研究では、場所表現を含むテキストの文脈情報を活用しているが、本研究では、文脈情報が不足する状況に着目し、ユーザとの対話による曖昧性の解消を試みた。

本研究は、画像の位置推定 (Image Geolocation) 研究とも密接に関わる。Weyand ら [4] は、画像のピク

セル情報をもとに写真の撮影場所を推定する課題に対し、深層学習を活用した手法を提案している。この手法では、地球表面を多層的な地理的セルに分割し、数百万枚のジオタグ付き画像を用いて訓練した深層学習モデルによって、画像の位置を特定する。ランドマークや自然地形などの視覚的特徴を活用することで、高精度な位置推定を実現し、従来の画像検索手法よりも優れた性能を示した。

場所に関する対話という観点では、地図課題対話に関する研究が深く関連する。川端ら [5] は、対話による地図情報の伝達の有効性を検討している。ここで地図課題対話コーパスとは、目標物と経路の描かれた地図を持つ話者が、目標物のみ描かれた地図を持つ話者に対しルートを教え、再現する会話が収録されたものである。このように、地図やナビゲーションに関連するタスクにおいて、対話を通じた情報伝達についての研究は従来から行われおり、本研究は、この課題のランドマークの曖昧性解消の部分に焦点を当てたサブタスクともみなせる。

曖昧性解消の研究は、一般的に、語義曖昧性解消 (WSD) をはじめとして、文脈を考慮して適切な選択をする古典的タスクであり、近年でも機械翻訳などで用いられている [6]. 本研究では、文脈がテキストでなく、画像で与えられ、語義は位置となる WSD の亜種ともみなせる。

## 6 おわりに

本研究では、曖昧な場所表現の位置を特定するためのユーザへの質問内容生成を目的として、複数の場所候補を絞り込む上で手がかりとなるランドマークの抽出を試みた。実験では、二つの店舗の識別に有用なランドマークを抽出するタスクにおいて、LLM の有用性を検証した。結果として、モデルは中程度の性能を示したが、カテゴリの曖昧さや過剰な推測に起因する正解ランドマークの見逃しが多い傾向にあった。今後の課題として、プロンプト設計の改善や、実際に抽出したランドマークをもとに生成した質問によるユーザへの検証が考えられる。

## 謝辞

本研究は、JST, RISTEX, JPMJRX20B2 の支援を受けたものです。

## 参考文献

- [1] Xuke Hu, Zhiyong Zhou, Hao Li, Yingjie Hu, Fuqiang Gu, Jens Kersten, Hongchao Fan, and Friederike Klan. Location reference recognition from texts: A survey and comparison. Vol. arXiv:2207.01683, , 2022.
- [2] Zheren Yan, Can Yang, Lei Hu, Jing Zhao, Liangcun Jiang, and Jianya Gong. The integration of linguistic and geospatial features using global context embedding for automated text geocoding. **ISPRS International Journal of Geo-Information**, Vol. 10, No. 9, pp. 1–14, 2021.
- [3] Gritta Milan, Pilehvar MT, and Collier Nigel. Which melbourne? augmenting geocoding with maps. **Association for Computational Linguistics**, pp. 1285–1296, 2018.
- [4] Tobias Weyand, Ilya Kostrikov, and James Philbin. Planet - photo geolocation with convolutional neural networks. **Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VIII 14**, pp. 37–55, 2016.
- [5] 川端良子, 松香敏彦, 伝康晴. 地図課題対話において発話の理解はどのように示されるか? 日本認知科学会第33回大会発表論文集, pp. 820–824, 2016.
- [6] Niccolò Campolungo, Federico Martelli, Francesco Saina, and Roberto Navigli. DiBiMT: A novel benchmark for measuring Word Sense Disambiguation biases in Machine Translation. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 4331–4352, Dublin, Ireland, May 2022. Association for Computational Linguistics.